DOCUMENT RESUME

ED 272 521 TM 860 426

AUTHOR Mershon, Clarence
TITLE How Parkrose Improved Its Testing Program Using the Rasch Model.
PUB DATE 17 Apr 86
NOTE 27p.; Paper presented at the symposium, "Taming the Rasch Tiger: Using Item Response Theory in Practical Educational Measurement," at the Annual Meeting of the National Council on Measurement in Education (San Francisco, CA, April 17-20, 1986). For other symposium papers, see TM 860 424-428.
PUB TYPE Viewpoints (120) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; *Criterion Referenced Tests; Elementary Secondary Education; Graduation Requirements; Graphs; *Latent Trait Theory; Mathematical Models; *Standardized Tests; *Testing Programs; *Test Theory
IDENTIFIERS Calibration; *Parkrose School District OR; *Rasch Model

ABSTRACT

In 1972, the Oregon State Department of Education changed its standards relating to graduation requirements after a survey revealed concern that the high school diploma had lost its credibility. Parkrose School District developed a goal-based curriculum and measurement program to meet the new requirements. From the program's inception, District leadership perceived the state mandate as an opportunity to re-define curriculum in certain critical areas and to develop a program to monitor student performance in those areas. Coincidentally with goal-based curriculum definition, criterion referenced tests were developed to assess student performance in reading, language, mathematics and science. As these developments progressed, issues surfaced that required attention and resolution, including: (1) district curriculum vs. "teacher" curriculum; (2) test items vs. curriculum; (3) number of items for goal coverage; (4) timeliness in reporting results; (5) test security; and (6) standards. Tests were developed to measure achievement with respect to defined learner outcomes, and development teams wrote test items referencing specific goals. The District established a performance standard at 80% for most skill areas. Reporting and scoring time was reduced, and a testing coordinator administered the tests to assure security. The availability of computer software allowed for a student database, and provided the capability of applying Rasch and other test analysis techniques. Appendices include graphic profiles of standard performance trends. (JAZ)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED272521

How Parkrose Improved Its Testing Program Using The Rasch Model

by

Clarence Mershon, Principal
Director Of Program Evaluation
Parkrose School District
10636 N.E. Prescott Street
Portland, Oregon 97220

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

C. Mershon

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the National Council for Measurement in Education

San Francisco, California
April 17, 1986

TM 860 426

How Parkrose Improved Its Testing Program Using The Rasch Model
By Clarence Mershon

My presentation will cover the development and refinement of the testing program in Parkrose School District since 1973. In particular, I will present data showing the way in which we are using item banking and the associated curriculum scales to assess and improve our programs.

Parkrose School Distict's goal-based curriculum and measurement program developed as a result of a change in Oregon State Department of Education standards relating to graduation requirements. In 1969, the Department surveyed Oregon's educational needs by sampling opinion of students and dropouts, educators and the public. The survey revealed widespread concern that the high school diploma had lost its credibility and that not all students were receiving training and instruction necessary to cope with the demands of society. The first draft of the resulting new graduation requirements, based upon student outcomes, was distributed in the fall of 1971. After review by interested individuals and groups, the new standards were adopted in 1972, and were to affect the graduating class of 1978.

In Parkrose, a steering committee was formed in 1972, and developed plans and an implementation time-line. In the fall of 1973, a Project Director was appointed to manage the effort. Planning involved administrators, teachers K-12, students and parents. Developmental teams were formed and were given responsibility for the curriculum planning at the program level. As the general framework for meeting State requirements emerged, sub-committees were formed in instructional program areas (such as mathematics, reading, science, etc.).

From the program's inception, District leadership perceived the State mandate as an opportunity to re-define curriculum in certain critical areas and to develop a program to monitor student performance in those areas. Coincidentally with goal-based curriculum definition, criterion referenced tests were developed to assess student performance in reading, language, mathematics and science. As these developments progressed, issues surfaced that required attention and resolution:

(1) District Curriculum vs. "Teacher" Curriculums: This issue relates to the fact that a district adopted curriculum does not insure its use in the classroom.

(2) Items vs. Curriculum: This issue concerns the alignment of test items to the defined curriculum. Do the test items used measure performance vis a vis a particular goal?

(3) Number of Items for Goal Coverage: How many test items are needed to give reliable information about performance in relationship to a particular goal?

(4) Timeliness in Reporting Results: In Parkrose this seemed to be a problem. With standardized tests, scoring and reporting took three weeks or more. We decided to reduce the time for scoring/reporting test results.

(5) Test Security: What steps can be taken to eliminate teaching to a test?

(6) Standards: What is an appropriate way to establish criteria for a "passing" score? How does one ensure that standards are not too easy or too difficult? How can one be certain that parallel test forms are of a near equal difficulty level?

Each of these problems was addressed as development proceeded. Their resolution, I believe, came about because District personnel were pragmatic, innovative in many respects and responsive to concerns of our constituency - teachers, students and parents.

It quickly became apparent that the District needed to do more than program definition. A District "curriculum" was extant, but teachers obviously did not feel bound to follow it. It established broad, general goals with neither specificity in outcomes nor direct measurement of achievement with respect to that curriculum. Accountability for results did not exist except at a most general level. Fortunately, the State mandate provided the impetus needed to help overcome reluctance to make the system more accountable. Starting in 1973, the curriculum was defined in terms of goal statements, a practical, intermediate level between the prior broad, generalized curriculum statements and what we perceived to be a minutiae of very specific behavioral objectives defined by some districts. In order to focus attention upon these goals, tests were developed, where appropriate, to measure achievement with respect to defined learner outcomes.

The next problem involved finding test items which were referenced to the more specific goals established. At that time, a course goal collection published by our local Educational Service District (Multnomah Educational Service District - M.E.S.D.) was available, but item collections referenced to a system of goals did not exist in useable form. Therefore, District development teams were paid to write items for specific goals. A special project in 1974, under the auspices of the Oregon Association for Supervision and Curriculum Development, provided an opportunity for staff members to be trained in writing items. High School classes graduating under the prior standards (classes of 1976 & 1977) were the subjects for field testing these items. Decisions were then required concerning the number of items per goal, the number of goals to be tested, the criteria for "passing" and similar issues.

Originally, many teachers believed that the standard should be 100%, but results from the field tests quickly demonstrated the unreasonableness of such a standard. Also, the field tests helped us in eliminating redundancies (where one skill is dependent upon another). The District established the performance standard at 80% for most skill areas, which implied some multiple of five test items for each goal area. Since reliability increases with the number of items, we established a minimum test length of twenty items, whether measuring one or more goals. Information available, together with our own follow-up studies, convinced us that this minimum (20 items) is necessary. The availability of the Rasch scaling system provides us with another possible standard for "passing" (though we do not use it at this point).

One factor negatively affecting the use of test results in Parkrose prior to this project was the time span between testing students and receiving test results from M.E.S.D. Generally this was three weeks or more. Also, the ability of teachers to analyze, synthesize and/or use the results for evaluative purposes was limited. Many questioned the content validity of standardized tests as well. The latter argument, of course, did not apply to our tests. To deal with the timeliness issue, a decision was made to score and report test results back to the teacher and student within twenty-four hours. This commitment enabled teachers to use the results immediately. In retrospect, establishing this policy of quickly providing test data to teachers certainly helped convince them of the usefulness of test results. Teacher reports concerning the impact upon student motivation were universally positive. Also, teachers were most appreciative of our practice of maintaining all performance records, rather than requiring them to do so.

Another issue that arose very soon was the problem of test security. The District made a decision to have a testing coordinator administer the tests. Neither teachers nor students were to have access to the test prior to the time of administration. However, during the initial year the test coordinator reported previous tests posted in some rooms, and instances were reported of teachers using prior tests as a teaching tool. As a consequence, the Director started the development of an item collection with specific reference to District goal areas. Concomitantly, a system for utilizing a bank to publish print ready copy was created. Shortly thereafter, the Northwest Evaluation Association (N.W.E.A.), a consortium of Pacific Northwest School Districts, came into being to provide a forum for test development, field testing and test research. As a participant in that effort, Parkrose contributed test items which were field tested and calibrated using responses from students throughout the states of California, Michigan, Oregon and Washington, and which were added to the N.W.E.A. item collections. Because of the possibility of being able to construct parallel test forms at a given level of difficulty, Parkrose School District was eager to cooperate fully in this developmental work. Use of the collections and our ability to calibrate new items makes it possible to develop District tests covering the desired content at the desired level of difficulty.

This latter issue has provided the foci for our efforts the past few years. Over the past twelve years, our testing program has changed from a mostly "hand" to a mostly "automated" system. The availability of microprocessor software which provides a student data base, and which gives us the capability of applying Rasch and other test analyses techniques enhances our work in this area. Initially our goal-based testing program focused upon graduation competence. As this program started to impact student achievement scores, the District decided to expand the program to include all levels. Changes in State standards in 1976 gave impetus to this effort, since these new standards required Oregon districts to monitor student performance in basic skill areas (reading, language and mathematics), K-12.

Our student data base provides the flexibility to test one time during the year and use information throughout the year. For example, data from achievement testing in the spring(Tests of Individual Performance - T.I.P.) are used to produce reports that match fall classes. Teachers have information day one to begin planning before students show up in the fall. The T.I.P. student data base enables the District to maintain records and monitor student growth from year to year, grade by grade. We use the data base to calculate graphic profiles of achievement growth for individual students and groups of students from year to year (See Appendices A and B). These data provide the foundation for the evaluation of school and District instructional programs. Results from goal-based and competence tests are similarly compiled and reported as needed.

The data base also provides an effective method for calibrating new test questions needed to assess District goals. We use the Fixed Parameter method to match the T.L.P. results to District developed tests and calibrate the items to the N.W.E.A. scales in reading, mathematics and language usage.

This past spring (1985) student responses on District developed tests and the T.L.P. data were used to calibrate many of our District tests. Appendix C and D show an example of the fixed parameter calibration and scaling for the third and fifth grade reading test, respectively. Similar analyses were performed on the third grade mathematics test, the sixth grade mathematics test and a mathematics competency examination.

District personnel have been concerned that some of our tests were either too easy or too difficult. The Director of Program Evaluation predicted that both third grade tests (reading and mathematics) were too easy and the fifth grade reading test too difficult. Table 1 contains the comparisons of the average RIT level of students as measured by the T.L.P. tests and the District's 80% standard established for each of these tests. As shown by the data, these three predictions were confirmed by results from the Fixed Parameter analysis.

Table 1

District Goal-Based Test	Rasch Level-80% Standard	District Average Rasch Level
Third grade reading	190	194.9
Third grade mathematics	187	193.8
Fifth grade reading	211	208.2
Sixth grade mathematics	222	223.1
*Test 1, Competencey (Graduation) Examination	229	230.6

*This test is given only to those 6th grade students who meet all criteria on the 6th grade mathematics test.

This research has made it possible to modify these tests to ensure that the 80% performance standard accurately reflects the achievement level desired. In the future, items from these tests can be used together with those in the item banks to construct parallel forms, expand the item collections and improve all tests.

Parkrose faces a dilemma similar to that experienced by other districts trying to improve its assessment program. Teachers are rightly concerned that tests measure performance vis a vis the curriculum. On the other hand, a district cannot long operate in isolation from what is occurring in other districts, hence the need for standardized measurement. Developing a system to accomplish the former without jeopardizing the latter is essentially what Parkrose is about. We also believe that student achievement scores will improve as the measurement and reporting systems improve, particularly as the District is able to establish more appropriate performance standards.

APPENCICES

A 1 to A 3

B 1 to B 6

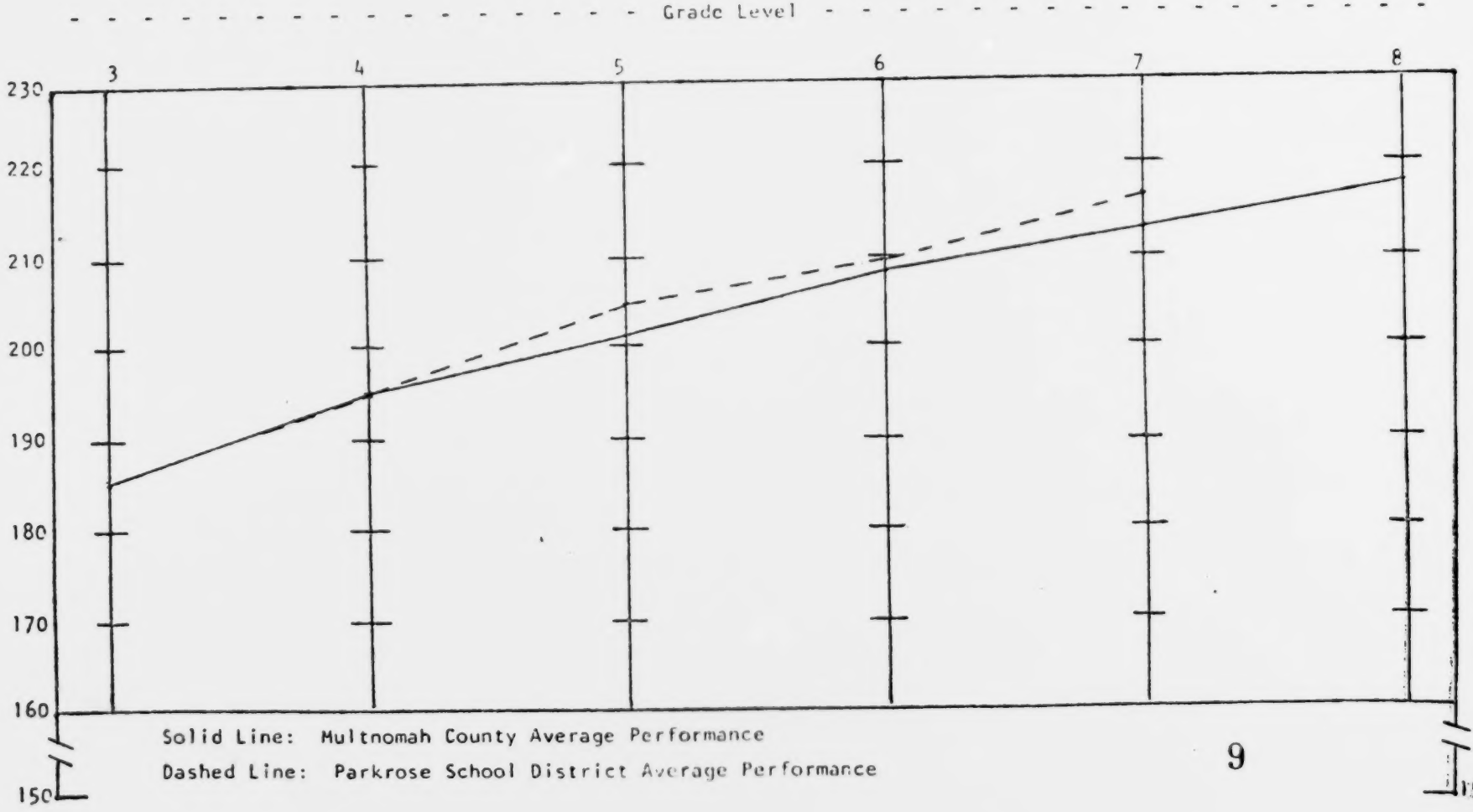
C 1

D 1

STANDARD SCORE PERFORMANCE TRENDS

School District
Class 1990
Test Reading (Fall)

PARKROSE SCHOOL DISTRICT



A 1

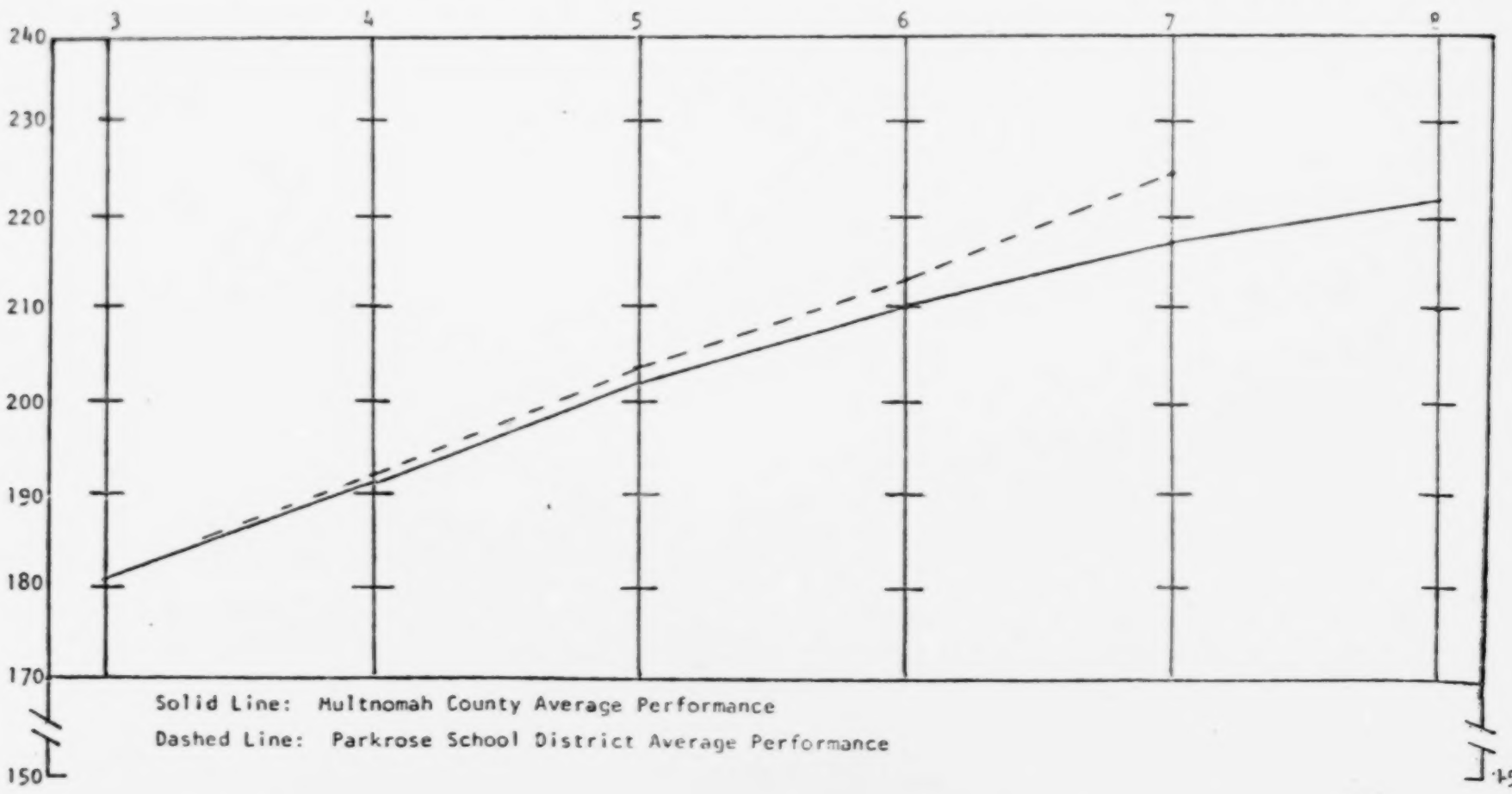
8

9

STANDARD SCORE PERFORMANCE TRENDS

School District
Class 1990
Test Mathematics (Fall)

PARKROSE SCHOOL DISTRICT

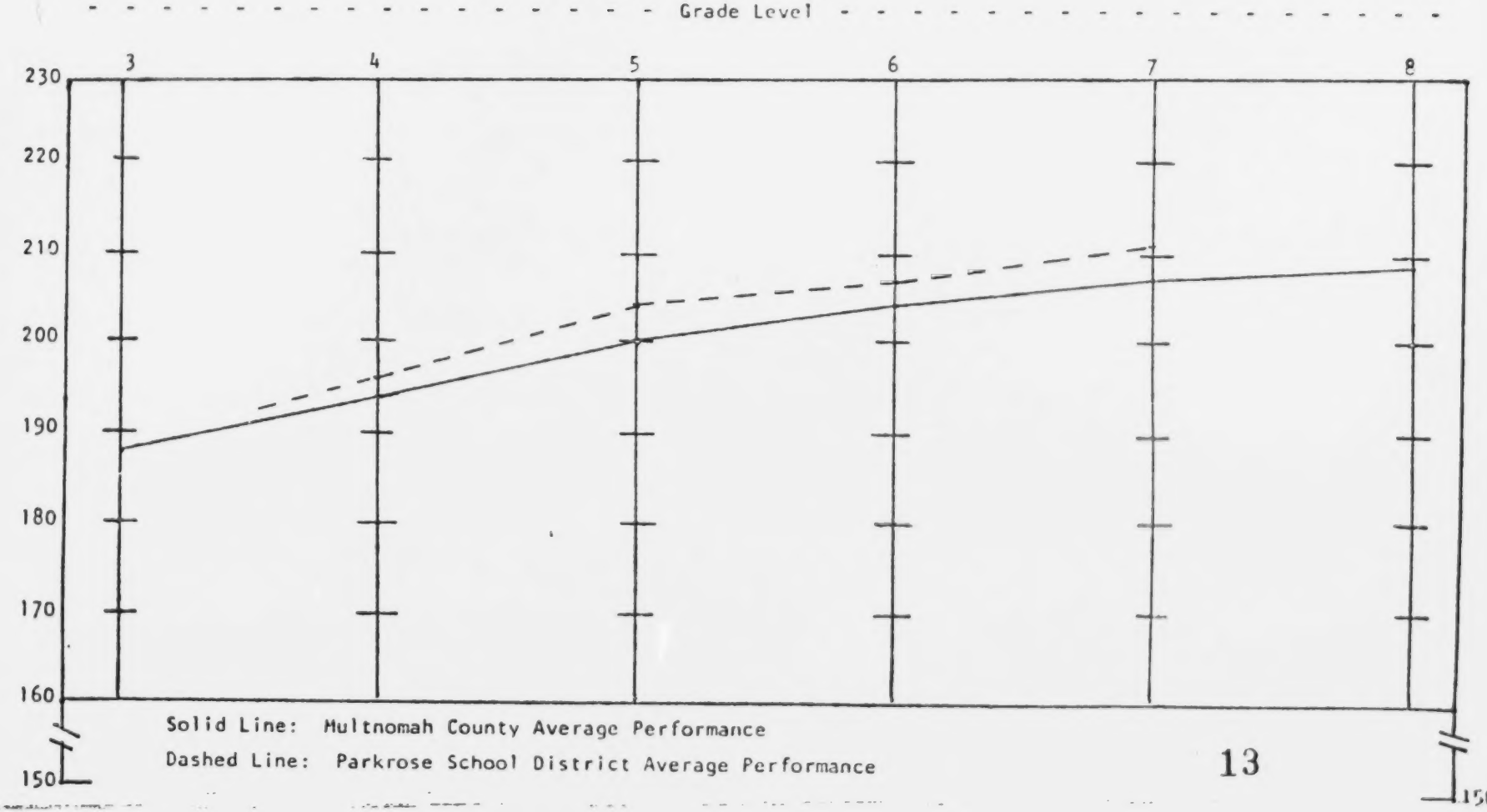


A 2

STANDARD SCORE PERFORMANCE TRENDS

School District
Class 1990
Test Language (Fall)

PARKROSE SCHOOL DISTRICT

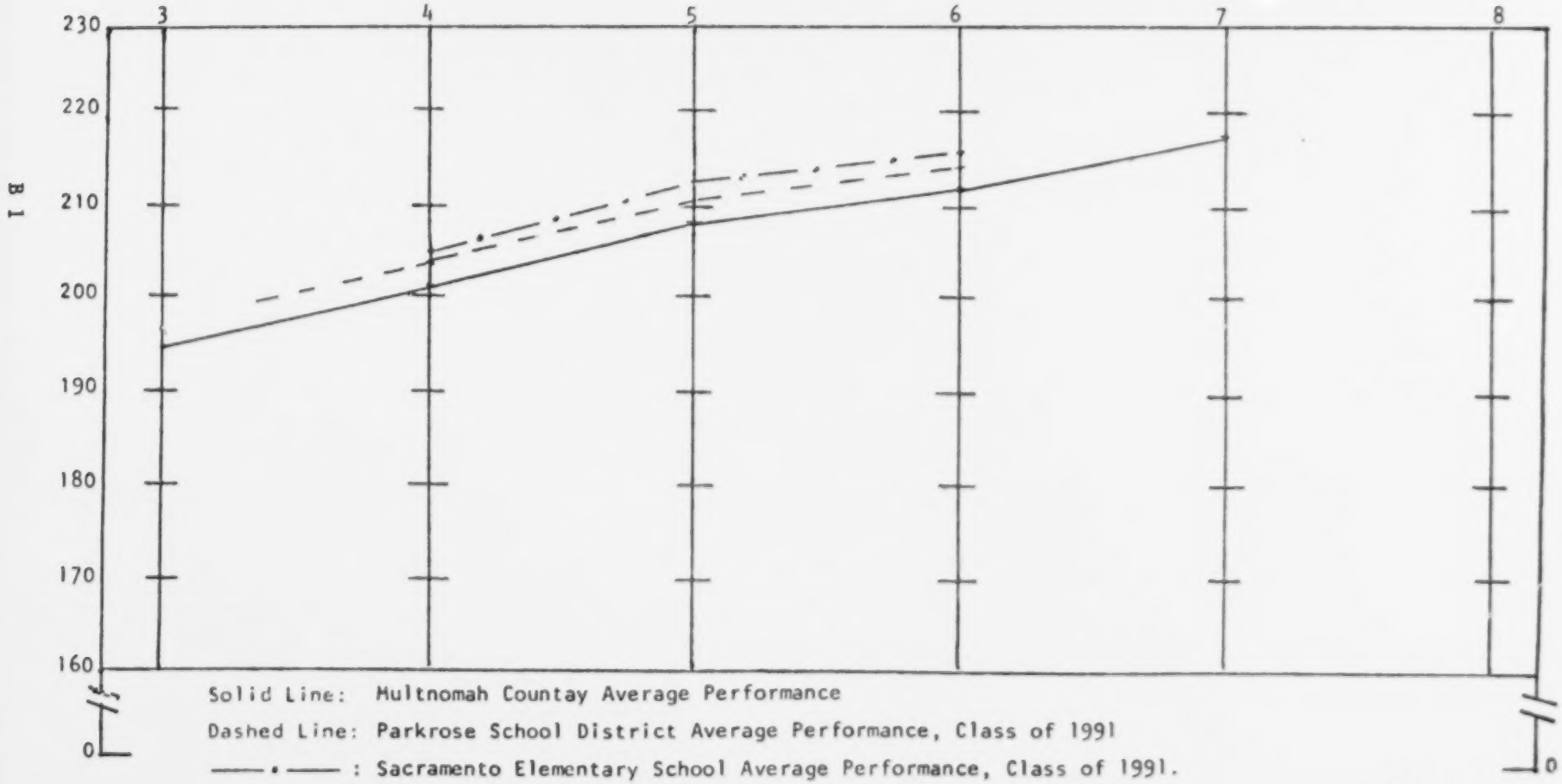


A 3

STANDARD SCORE PERFORMANCE TRENDS

School Sacramento
Class 1991
Test T.I.P. Reading (Spring)

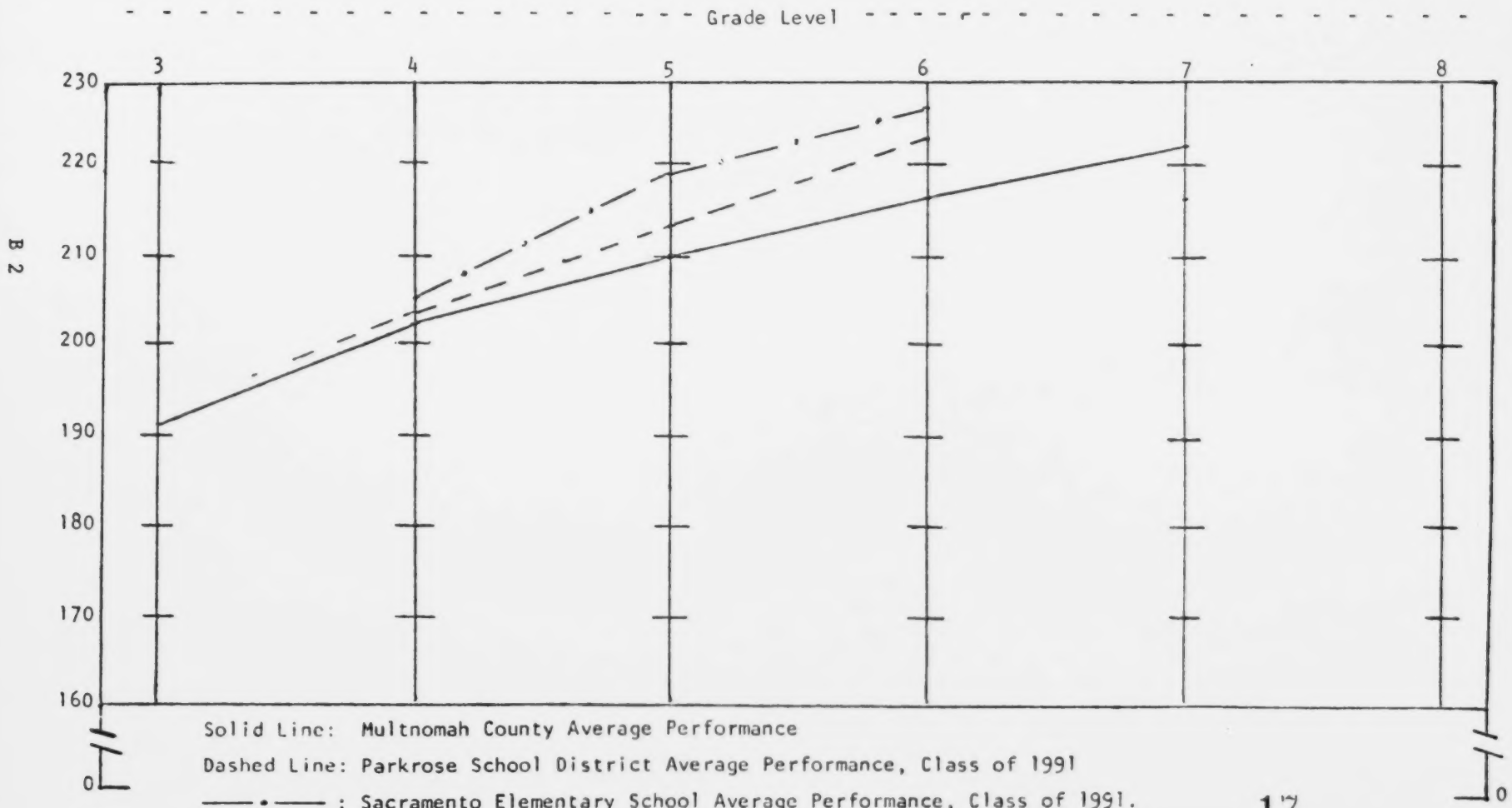
PARKROSE SCHOOL DISTRICT



STANDARD SCORE PERFORMANCE TRENDS

School Sacramento
Class 1991
Test T.I.P. Mathematics (Spring)

PARKROSE SCHOOL DISTRICT



B 2

Solid Line: Multnomah County Average Performance

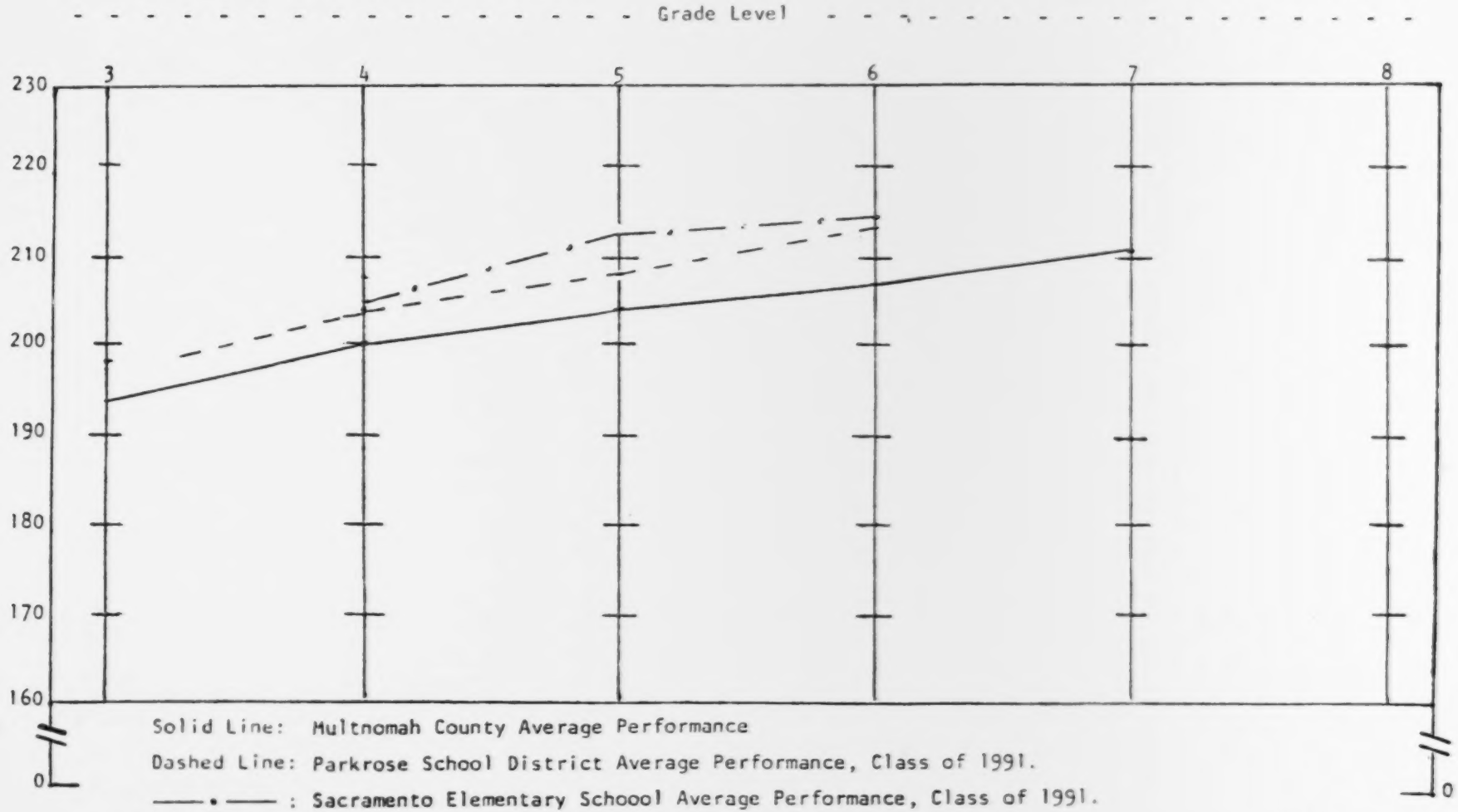
Dashed Line: Parkrose School District Average Performance, Class of 1991

——•—— : Sacramento Elementary School Average Performance, Class of 1991.

STANDARD SCORE PERFORMANCE TRENDS

School Sacramento
Class 1991
Test T.I.P. Language (Spring)

PARKROSE SCHOOL DISTRICT



Solid Line: Multnomah County Average Performance

Dashed Line: Parkrose School District Average Performance, Class of 1991.

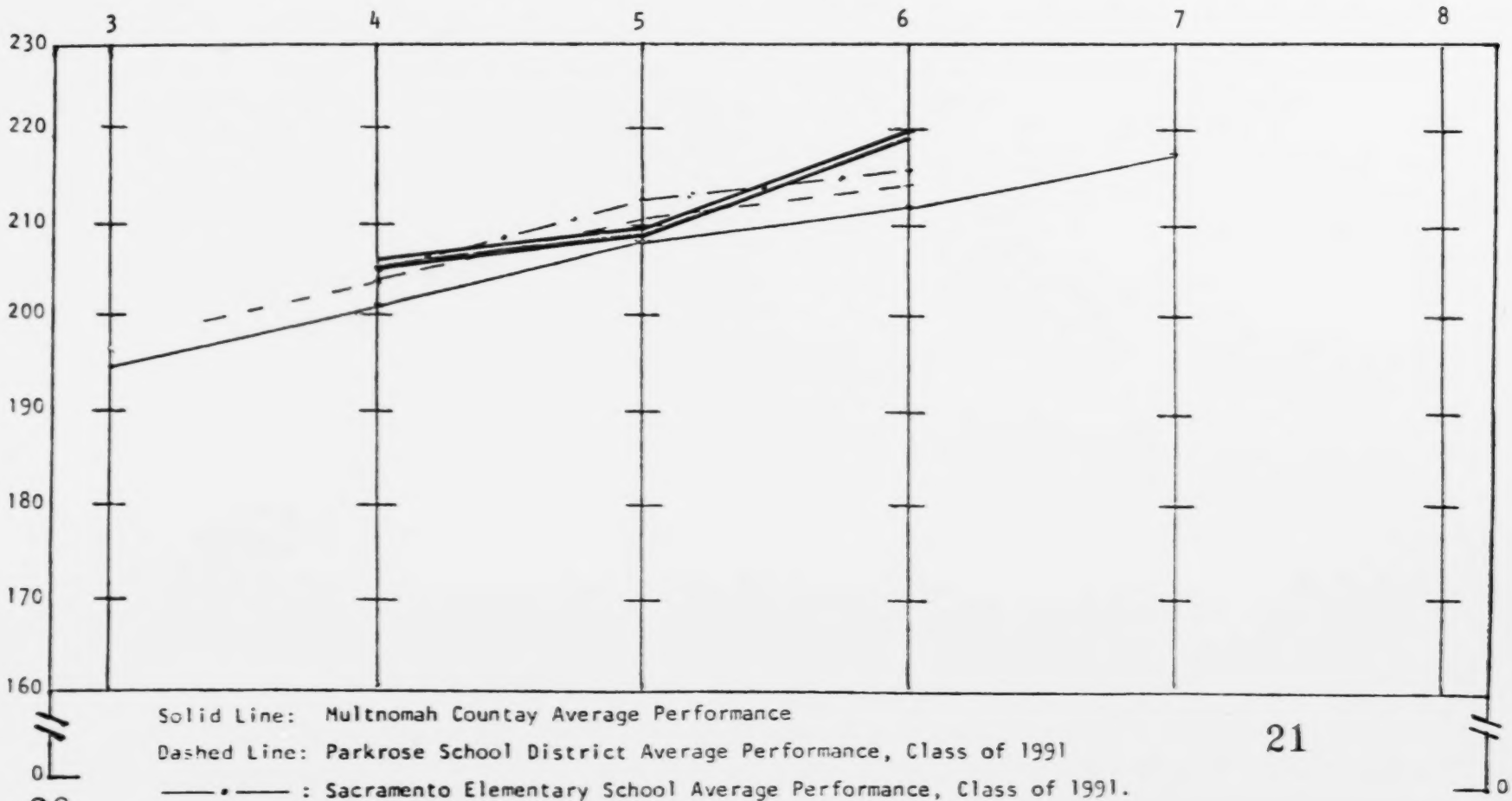
——•—— : Sacramento Elementary Schoool Average Performance, Class of 1991.

B 3

STANDARD SCORE PERFORMANCE TRENDS

School Sacramento
Class 1991
Test T.I.P. Reading (Spring)

PARKROSE SCHOOL DISTRICT



B 4

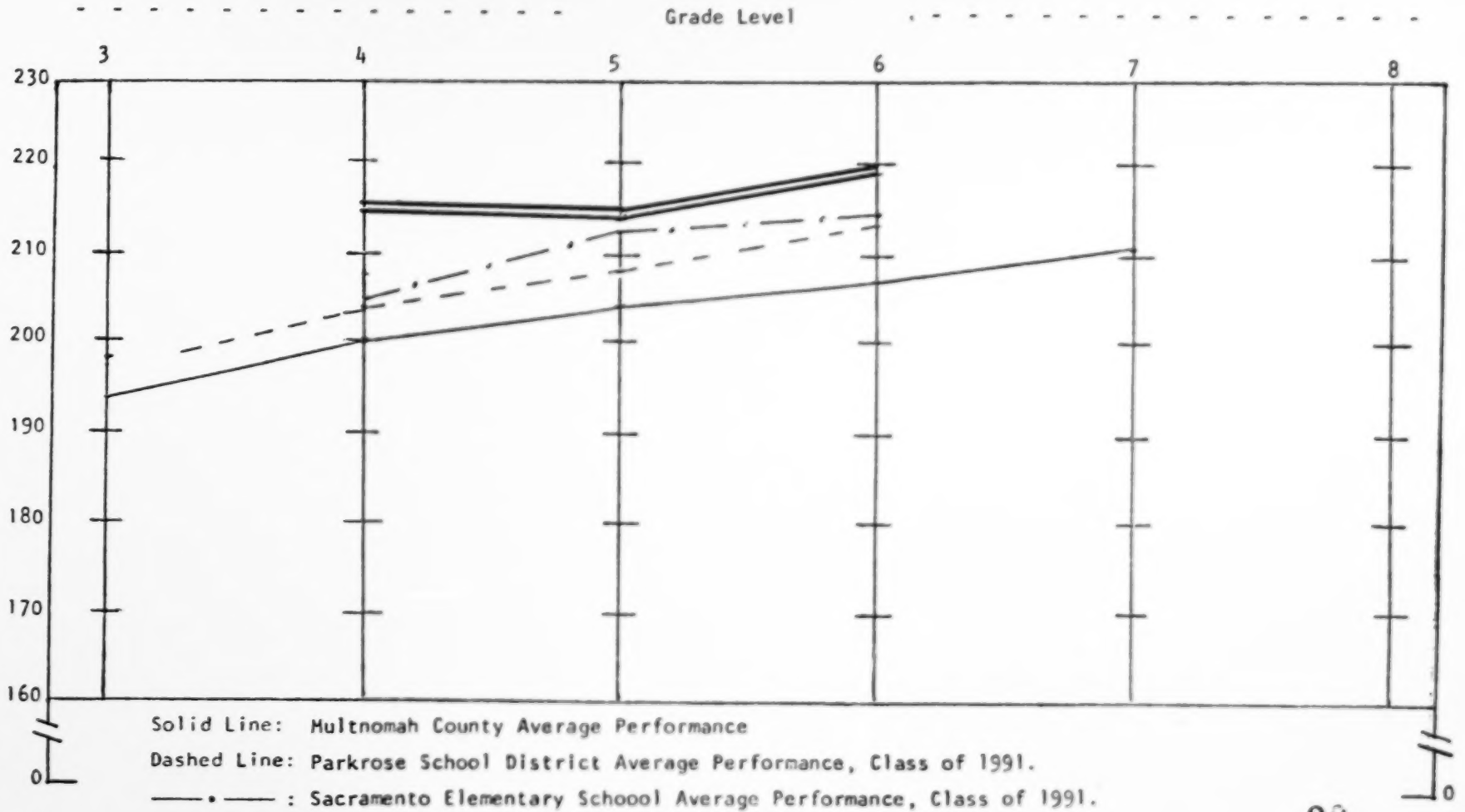
20

21

STANDARD SCORE PERFORMANCE TRENDS

School Sacramento
Class 1991
Test T.I.P. Language (Spring)

PARKROSE SCHOOL DISTRICT

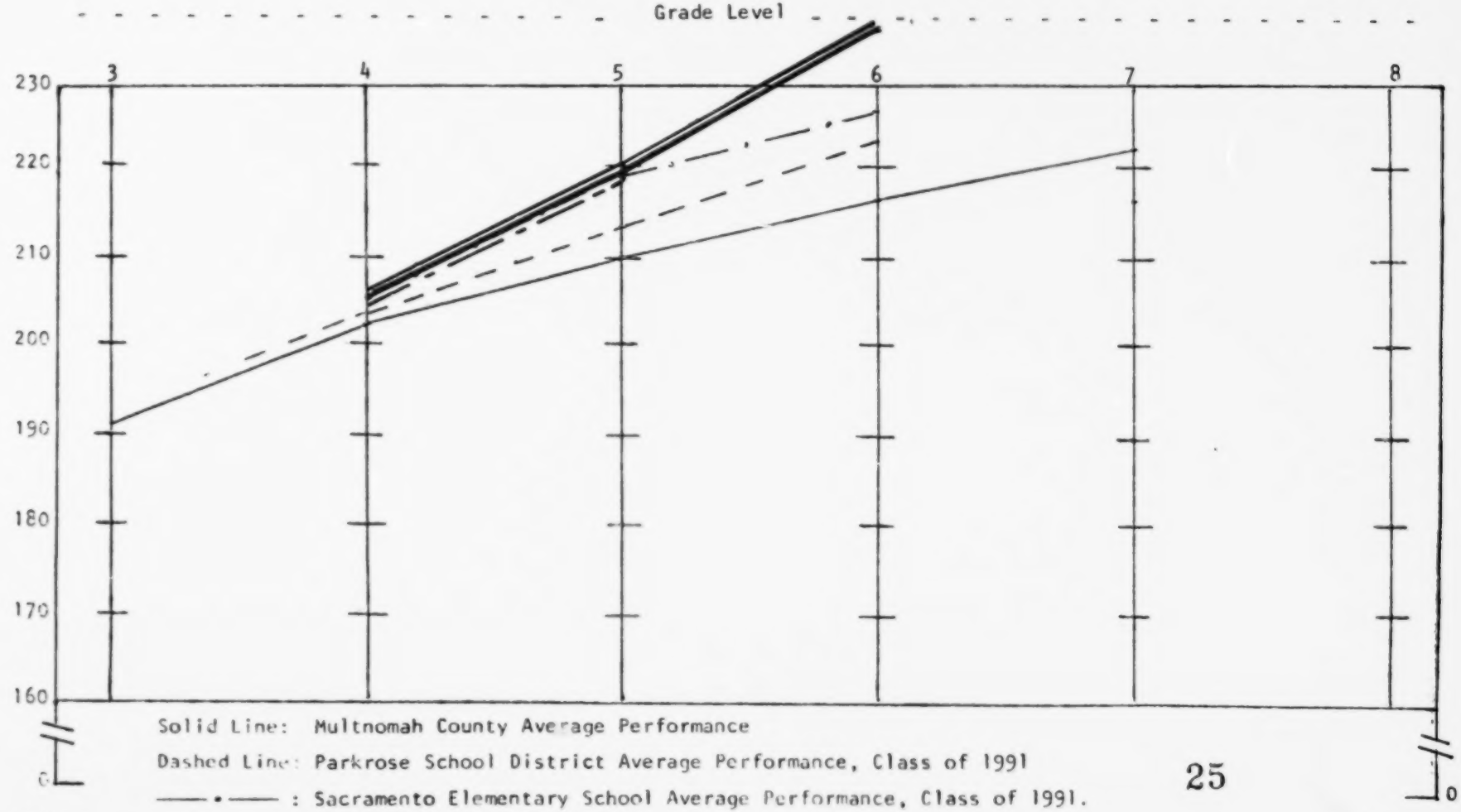


B 5

STANDARD SCORE PERFORMANCE TRENDS

School Sacramento
Class 1991
Test T.I.P. Mathematics (Spring)

PARKROSE SCHOOL DISTRICT



FIXED PARAMETER ANALYSIS REPORT FOR TEST 109 PARKROSE 3RD GRADE READING 02/27/86

219 STUDENTS TIP RASCH AVERAGE = 194.954 STD DEV = 15.1378 CRT RAW SCORE AVERAGE = 35.8767 STD DEV = 8.24501

ITEM	ITEMID	KEY	PERCENT	PTBIS	ITEM-RIT	GOAL	TEST-RIT	ERROR	GOAL	RIT
1	001	1	95	0.52	152	1	134	10.2	GOAL 1 -- 15 ITEMS	
2	002	2	90	0.21	164	1	142	7.4	1	142
3	003	3	84	0.33	171	1	146	6.1	2	150
4	004	3	87	0.35	168	1	150	5.4	3	156
5	005	2	79	0.52	176	1	152	4.9	4	160
6	006	1	84	0.43	172	1	154	4.6	5	163
7	007	3	76	0.48	179	1	156	4.3	6	166
8	008	1	89	0.29	166	1	158	4.1	7	169
9	009	3	79	0.51	176	1	160	3.9	8	172
10	010	2	86	0.58	170	1	161	3.8	9	175
11	011	1	81	0.55	174	1	163	3.6	10	178
12	012	3	75	0.46	179	1	164	3.5	11	182
13	013	4	89	0.52	166	1	165	3.5	12	186
14	014	2	85	0.61	170	1	166	3.4	13	191
15	015	2	80	0.63	176	1	167	3.3	14	199
16	016	2	73	0.55	181	2	168	3.3	GOAL 2 -- 15 ITEMS	
17	017	1	83	0.57	173	2	169	3.2	1	147
18	018	1	89	0.42	165	2	171	3.2	2	155
19	019	2	84	0.57	171	2	172	3.2	3	160
20	020	4	71	0.40	182	2	173	3.2	4	165
21	021	3	84	0.57	172	2	174	3.2	5	168
22	022	2	86	0.49	170	2	175	3.2	6	171
23	023	2	72	0.45	182	2	176	3.2	7	174
24	024	3	83	0.33	173	2	177	3.2	8	177
25	025	1	88	0.46	167	2	178	3.2	9	180
26	026	3	82	0.53	173	2	179	3.2	10	184
27	027	4	83	0.54	173	2	180	3.2	11	187
28	028	2	49	0.20	196	2	181	3.3	12	192
29	029	3	81	0.54	175	2	182	3.3	13	197
30	030	2	63	0.31	188	2	183	3.3	14	206
31	031	3	80	0.52	176	3	184	3.4	GOAL 3 -- 15 ITEMS	
32	032	4	81	0.31	174	3	185	3.5	1	151
33	033	4	89	0.54	166	3	186	3.5	2	159
34	034	2	80	0.42	176	3	188	3.6	3	164
35	035	1	76	0.50	179	3	189	3.8	4	168
36	036	4	82	0.41	173	3	190	3.9	5	171
37	037	3	76	0.50	179	3	192	4.1	6	174
38	038	3	61	0.46	189	3	194	4.3	7	177
39	039	3	69	0.52	184	3	196	4.6	8	180
40	040	4	77	0.55	178	3	198	4.9	9	183
41	041	4	80	0.49	176	3	201	5.4	10	186
42	042	1	76	0.53	179	3	204	6.1	11	189
43	043	3	62	0.39	188	3	208	7.4	12	193
44	044	4	73	0.65	181	3	216	10.2	13	198
45	045	1	73	0.54	181	3			14	206

FIXED PARAMETER ANALYSIS REPORT FOR TEST 101 PARKROSE 5TH GRADE READING 02/27/86

224 STUDENTS TIP RASCH AVERAGE = 208.665 STD DEV = 12.8682 CRT RAW SCORE AVERAGE = 28.9464 STD DEV = 7.1718

ITEM	ITEMID	KEY	PERCENT	PTBIS	ITEM-RIT	GOAL	TEST-RIT	ERROR	GOAL	RIT
1	001	1	79	0.35	191	1	156	10.2	GOAL 1 -- 10 ITEMS	
2	002	3	72	0.48	196	1	164	7.4	1	172
3	003	3	62	0.44	203	1	168	6.1	2	180
4	004	1	54	0.25	207	1	172	5.4	3	186
5	005	2	76	0.40	193	1	174	4.9	4	191
6	006	3	77	0.50	193	1	176	4.6	5	195
7	007	3	87	0.43	183	1	178	4.3	6	200
8	008	1	70	0.10	198	1	180	4.1	7	205
9	009	3	64	0.52	201	1	182	4.0	8	211
10	010	4	82	0.57	189	1	183	3.8	9	219
11	011	4	68	0.52	199	2	185	3.7	GOAL 2 -- 10 ITEMS	
12	012	2	74	0.58	195	2	186	3.6	1	169
13	013	4	77	0.47	193	2	188	3.6	2	178
14	014	2	88	0.36	190	2	189	3.5	3	184
15	015	3	72	0.56	196	2	190	3.5	4	189
16	016	3	89	0.55	181	2	191	3.4	5	194
17	017	1	91	0.29	178	2	192	3.4	6	198
18	018	2	61	0.29	203	2	193	3.4	7	203
19	019	4	58	0.21	205	2	195	3.3	8	209
20	020	3	76	0.31	194	2	196	3.3	9	218
21	021	2	81	0.34	189	3	197	3.3	GOAL 3 -- 5 ITEMS	
22	022	1	58	0.29	205	3	198	3.4	1	179
23	023	4	87	0.54	184	3	199	3.4	2	190
24	024	4	66	0.48	200	3	200	3.4	3	202
25	025	4	71	0.45	197	3	201	3.4	4	211
26	026	2	73	0.36	196	4	203	3.5	GOAL 4 -- 5 ITEMS	
27	027	1	76	0.38	194	4	204	3.5	1	178
28	028	4	84	0.43	187	4	205	3.6	2	188
29	029	1	83	0.45	188	4	206	3.7	3	197
30	030	2	70	0.45	198	4	208	3.8	4	207
31	031	4	60	0.52	203	5	209	3.9	GOAL 5 -- 5 ITEMS	
32	032	2	56	0.42	206	5	211	4.1	1	185
33	033	4	66	0.43	200	5	213	4.3	2	196
34	034	2	70	0.44	198	5	215	4.5	3	204
35	035	2	78	0.41	192	5	217	4.9	4	214
36	036	3	78	0.48	192	6	219	5.4	GOAL 6 -- 5 ITEMS	
37	037	4	53	0.44	207	6	223	6.1	1	185
38	038	2	72	0.59	196	6	227	7.3	2	195
39	039	3	58	0.27	205	6	234	10.2	3	204
40	040	3	69	0.42	198	6			4	214